On the Notion of Social Institutions*

Guilherme Carmona
Universidade Nova de Lisboa

October 15, 2002

Abstract

We argue that it is natural to study social institutions within the framework of standard game theory (i.e., only by resorting to concepts like players, actions, strategies, information sets, payoff functions, and stochastic processes describing the moves of nature, which constitute a stochastic game when combined) — concepts like "social norms," and "mechanisms" can be easily accommodated, as well as philosophical/sociological definitions of social institutions

Focusing on strategies rather than on mechanisms have two advantages: First, focusing on strategies allows us to distinguish between those aspects that are behavioral in nature and are subject to alternative design, and those that are part of the environment. Second, considering strategies allows for a more detailed look into the way an outcome function is "genuinely implemented" (Hurwicz (1996, p. 123)).

*This paper corresponds to Chapter 2 of my Ph.D. thesis. I wish to thank my advisors, Narayana Kocherlakota, and Marcel K. Richter, for their guidance, encouragement, and very helpful comments and suggestions. I am also very grateful to the other members of my dissertation committee, Leonid Hurwicz, and Max Jodeit. While they all improved the material in many ways, all remaining errors are mine. This paper tries to address some questions posed to me by Leo Hurwicz, who, without being in complete agreement with some of my answers, made invaluable comments. Financial support from the Subprograma Ciência e Tecnologia do 2º Quadro Comunitário de Apoio is greatufully acknowledged.

1 Introduction

Economists have long recognized the importance of social institutions in economic processes. However, attempts to formalize them have been relatively recent, and seemingly confliting. In particular, social institutions have been defined as strategies in a particular repeated game by Schotter (1981, p. 155), and Okuno-Fujiwara and Postlewaite (1995, p. 83), and also as families of game forms by Hurwicz (1996, p. 117). We will use a game-theoretic approach and define a social institution as a family of strategies. We then argue that Hurwicz's formulation can be encompassed in the game-theoretic formulation we use in a natural way, and discuss the advantages of doing so.

In the game-theoretic approach of Schotter (1981) and Okuno-Fujiwara and Postlewaite (1995) the first step consists of the description of the society we are interested in. A society will be described by listing all its members, all the actions that each one of the society members can take, and, finally, each member's preferences over action combinations. Hence, we naturally describe any society by a normal formal game. Further, we assume that the society will last forever, and that its members, their actions, and preferences can change through time. Thus, although a normal form game describes the society in a given point in time, its complete description is provided by a stochastic game.[1]

As in Hurwicz (1994, p. 6), we view *the space of outcomes* as the natural space over which society members have their preferences. Therefore we add a *physical outcome function* , to the stochastic game describing a given society, which gives the physical outcome of any action combination, and an *utility function* for each member — each member's payoff function is then the composition of the physical outcome function with that member's utility function. This formalization tries to capture a situation in which players' actions, together with a given realization of the uncertainty, produce an outcome according to the laws of physics, in a way that is described by the physical outcome function.

The stochastic game and the physical outcome function describe the (eco-

[1]It is important to emphasize the assumption that every person in the society is a player, and that any action that a person can take belongs to that person's action space. In some cases, the action spaces may not include all the actions that are physically possible but only all the actions that are of interest to us. For example, if we want to study whether it is possible to resolve some conflict without violence, then any action that involves violence is naturally excluded from the space of actions that players have available.

nomic) environment, and are considered to be given. In order to complete the description of a given society, we add the notion of strategies, which describe the behavioral aspects of society' members. A *strategy* is a set of rules that tell each member of the society how to act in any possible contingency. We then define a *social institution* as a family of strategies.

The following example tries to clarify our view. Consider a legal code in a given society. The effect of the legal code is that certain actions of certain people will be consider 'illegal,' and a certain punishment, to be enforced by a different group of people, will be associated with those actions. Thus, we associate a legal code with the set of all strategies respecting the above properties, i.e., all the strategies that will trigger the corresponding punishment whenever an illegal action is taken.

The example also shows why a social institution should be defined as a family of strategies rather than a particular strategy. If two strategies respect the defining properties of a legal code but differ on the legal actions a given player takes, then we would still say that the legal code prevails in both cases.

Our view is thus that social institutions are rules of behavior. Here is interesting to contrast our interpretation with that of North (1990, p. 3) and Schotter (1981, p. 155). For North, "[i]nstitutions are the rules of the game in a society or, more formally, are the humanly devised constraints that shape human interaction." Schotter has a different opinion:

> "For us, however, what we call social institutions are not the rules of the game but rather the alternative equilibrium standards of behavior or conventions of behavior that evolve from a given game described by its rules. In other words, for us, institutions are properties of the equilibrium of games and not properties of the game's description."

In our opinion, both views are acceptable depending on the focus we use: from the analyst's point of view, institutions are (families of) strategies; however, from the viewpoint of the players, institutions are rules that shape their interaction.[2]

[2]However, in contrast with Schotter (1981), we do not require that the outcome resulting from a given institution be an equilibrium outcome — hence, we allow in our notion of an institution all a priori possible institutions, although we are mainly interested in equilibrium institutions. This will allow us, in principle, to discuss which institutions are an equilibrium (and thus can be expected to endure) and those that are not.

Let us now describe Hurwicz's formulation (see Hurwicz (1994), and Hurwicz (1996).) In Hurwicz's formulation, a social institution is a family of mechanisms (synonymously, game forms). A mechanism consists of a strategy space for each player and an *outcome function*, mapping strategies into outcomes. Note that Hurwicz's outcome function differs from our physical outcome function because, while our physical outcome function reflects only physical aspects, "[q]uite often a game form has two aspects: (1) those that are behavioral in nature and are (potentially at least) subject to alternative design and embody the essence of institutional arrangements, and (2) those that are considered as given, either because they are determined by the laws of physics or, more generally, by at we call the (economic) environment (in particular by the existing resource endowments, and the current state of technology)." (Hurwicz (1996, p. 118).)

The objective of Hurwicz's outcome functions is to describe the consequences of players' choices of strategies. However, it is often enough to know how the outcome changes if only one player changes his strategy, as it is the case in equilibrium analysis. In this case, we only need, for every strategy and for every player, a mapping from that player's strategy space into the outcome space, which is a "partial" outcome function. Our point is that each strategy, together with the given physical outcome function, induces such a partial outcome function for every player, and thus, a family of strategies (i.e., a social institution) induces a family of partial outcome functions, which together with the given strategy space, define a family of mechanisms.

Hence, in our view, a game-theoretic formulation (as described above) does not conflict with the one proposed by Hurwicz. In fact, it can be viewed as an attempt to address the following two concerns expressed in Hurwicz (1996).

First, focusing on strategies allows us to separate explicitly the aspects that are behavioral in nature from those that are determined by the laws of physics or, more generally, by the environment. In our formulation, a given strategy is responsible for the behavioral aspects and the physical outcome function is responsible for the physical aspects. Since the behavioral aspects are the ones that are subject to alternative design and embody the essence of institutional arrangements, we associate social institutions with strategies.

Second, considering strategies instead of the (partial) outcome function induced by them allows for a more detailed look into the way an outcome function "works." As Hurwicz points out (Hurwicz (1996), page 123)

> "(...) the results specified by the outcome function must also be "delivered". Mechanisms such as markets or social insurance, illustrate the need for special machinery required to carry out, in addition to enforcement, the informational functions (in particular, communication) as well as the physical flow of goods and financial instruments. (...) [I] speak of "*genuine implementation*", with the intention of covering the complex of all activities designed to make the outcome function effective."

In other words, our approach provides a more detailed analysis of how an outcome function is genuinely implemented by analyzing the incentives of every person involved in it — including those responsible for enforcement, informational functions, and so on.

The argument above notwithstanding, there are applications where it may be more natural to work directly with outcome functions, instead of incorporating all the people that are involved in the problem, and all their possible actions. This is certainly the case when some aspects can be taken as given, or when modelling them explicitly would complicate the analysis without adding much insight. However, as we try to demonstrate in Chapter 4, our analysis is quite natural for certain problems.

Finally, we would like to acknowledge Professor Hurwicz's disagreement with the claim that the infinitely repeated game-theoretic formulation that we use is general enough to encompass the mechanism design formulation. Furthermore, in his opinion there is no contradiction between the two approaches because they reflect different models of emergence of social institutions: while the infinitely repeated game-theoretic approach tends to model social institutions that emerge endogenously, the mechanism design approach models their emergence as exogenous, or as a sequence of games played by different groups of players.

2 Notation and Definitions

A *noncooperative stochastic game* G is defined by

$$G = \left\langle \Omega,\ N, \langle A_i \rangle_{i\in N}, \left\langle v_\omega^i \right\rangle_{\omega\in\Omega,\ i\in N}\ , g\ , p \right\rangle,$$

where: (1) Ω, N and, for all $i \in N$, A_i are finite sets; (2) for all $\omega \in \Omega$, and $i \in N$, $v_\omega^i : A \to \mathbb{R}$, where $A = \prod_{i\in N} A_i$; (3) $g : \Omega \times A \to \Delta(\Omega)$; and (4) $p \in \Delta(\Omega)$.

It is convenient to define $G_\omega = \langle N, \{A_i\}_{i\in N}, \{v^i_\omega\}_{i\in N}\rangle$ for all $\omega \in \Omega$,; that is, G_ω is to be interpreted as a noncooperative normal form game. Then, the interpretation of G is as follows: there are denumerably many stages, in which some element of $\{G_\omega\}_{\omega\in\Omega}$ will be played. The probability that G_ω, $\omega \in \Omega$, is played in the first stage equals p_ω; the probability that G_j, $j \in \Omega$, is played in stage $k+1$, given that the stage games and the actions played from stage 1 up to stage k were $(G_{j_1}, ..., G_{j_{k-1}}, G_\omega, a_1, ..., a_{k-1}, a)$ equals $g(\omega, j)(a)$. We will typically use $g(a|\omega, j)$ for $g(\omega, j)(a)$.

Finally we need to specify the strategies players can use and also a way to evaluate payoffs in the supergame of G. For $k \geq 1$, a $k-$*stage history* is a $k-$length sequence $h_k = (\omega_1, a_1, \ldots, \omega_k, a_k)$, where, for all $1 \leq t \leq k$, $a_t \in A$; the space of all $k-$stage histories is H_k, i.e., $H_k = A^k$ (the $k-$fold Cartesian product of A.) The notation e stands for the unique 0–stage history — it is a 0–length history that represents the beginning of the supergame. The set of all histories is defined by $H = \bigcup_{n=0}^{\infty} H_n$.

The notation $\tilde{h}_k$ denotes (h_k, ω_{k+1}), with $h_k \in H_k$ and $\omega_k \in \Omega$; the space of all $\tilde{h}_k$ is denoted by $\tilde{H}_k$. It is assumed that at stage k each player knows $\tilde{h}_k$, i.e., each player knows the actions that were played in all previous stages, the states of nature of all previous stages and that of the current stage. As for strategies, in each stage k, every player $i \in N$ chooses a function $\sigma^i_k : H_{k-1} \to A_i$. The set of player i's strategies is denoted by Σ_i, and $\Sigma = \prod_{i\in N} \Sigma_i$ is the joint strategy space. Finally, a strategy vector is $\sigma = \left(\{\sigma^i_k\}_{k=1}^{\infty}\right)_{i\in N}$.

Let $k \in \mathbb{N}$, and $\sigma \in \Sigma$. For each $\omega^k = (\omega_1, \ldots, \omega_k) \in \Omega^k$, let $a_t(\omega^k)$, $1 \leq t \leq k$ be defined by induction as follows: $a_1(\omega^k) = \sigma(\omega_1)$ and $a_t(\omega^k) = \sigma(\omega_1, a_1(\omega^k), \ldots, \omega_{t-1}, a_{t-1}(\omega^k), \omega_t)$. The expected payoff in period k is then

$$v^i_k(\sigma) = \sum_{\omega^k\in\Omega^k} p(\omega_1)g(\omega_2|\omega_1, a_1(\omega^k)) \cdots g(\omega_k|\omega_{k-1}, a_{k-1}(\omega^k))v^i_{\omega_k}(a_k(\omega^k)).$$

The payoff in the stochastic game G is, for $\delta \in (0, 1)$, the expected discounted sum of stage game payoffs:

$$V^i(\sigma) = (1-\delta)\sum_{k=1}^{\infty} \delta^{k-1} v^i_k(\sigma).$$

For every $h \in H$, define $h^r \in \Omega \times A$ to be h's r^{th} coordinate. For every $h \in H$ we let $\ell(h)$ denote the *length of* h. For two positive length histories h and $\bar{h}$ in H we define the *concatenation of* h *and* $\bar{h}$, in that order, to be the

history $(h \cdot \bar{h})$ of length $\ell(h)+\ell(\bar{h})$: $(h \cdot h') = (h^1, h^2, \ldots, h^{\ell(h)}, \bar{h}^1, \bar{h}^2, \ldots, \bar{h}^{\ell(\bar{h})})$. We also make the convention that $e \cdot h = h \cdot e = h$ for every $h \in H$.

Similarly, for every $\tilde{h} \in H$, we let $\ell(\tilde{h}) = \ell(h)$ denote the *length of* $\tilde{h}$. For two positive length histories $\tilde{h}$ and $\tilde{h}'$ in $\tilde{H}$, with $\omega_{\ell(\tilde{h})+1} = \omega'_1$ we define the *concatenation of* $\tilde{h}$ *and* $\tilde{h}'$, in that order, to be the history $(\tilde{h} \cdot \tilde{h}')$ of length $\ell(\tilde{h}) + \ell(\tilde{h}')$:

$$\begin{aligned}(\tilde{h} \cdot \tilde{h}') &= (h^1, h^2, \ldots, h^{\ell(h)}, (\omega_{\ell(h)+1}, a'_1), h'^2, \ldots, h'^{\ell(\bar{h})}, \omega'_{\ell(h'+1)}) \\ &= (h^1, h^2, \ldots, h^{\ell(h)}, h'^1, h'^2, \ldots, h'^{\ell(h')}, \omega'_{\ell(h'+1)}).\end{aligned}$$

Given an individual strategy $\sigma_i \in \Sigma_i$ and a history $\tilde{h} \in \tilde{H}$ we denote the *individual strategy induced by* σ_i *at* $\tilde{h}$ by $\sigma_i|\tilde{h}$. This strategy is defined pointwise on $\tilde{H}$: $(\sigma_i|\tilde{h})(\tilde{h}') = \sigma_i(\tilde{h} \cdot \tilde{h}')$, for every $\tilde{h}' \in \tilde{H}$. We will use $(\sigma|\tilde{h})$ to denote $(\sigma_1|\tilde{h}, \ldots, \sigma_n|\tilde{h})$ for every $\sigma \in \Sigma$ and $\tilde{h} \in \tilde{H}$. We let $\Sigma_i(\sigma_i) = \{\sigma_i|\tilde{h} : \tilde{h} \in \tilde{H}\}$ and $\Sigma(\sigma) = \{\sigma|\tilde{h} : \tilde{h} \in \tilde{H}\}$.

A strategy vector $\sigma \in \Sigma$ is a *Nash equilibrium* of G if $V^i(\sigma) \geq V^i(\hat{\sigma}_i, \sigma_{-i})$ for all $\hat{\sigma}_i \in \Sigma_i$. A strategy vector $\sigma \in \Sigma$ is a *subgame perfect equilibrium* of G if every $\bar{\sigma} \in \Sigma(\sigma)$ is a Nash equilibrium.

In many circumstances, we are interested in the physical outcomes that are induced by the choices of the players. Hence, we assume that there is a set Z, referred to as the outcome space, and a function $P : A \to Z$, which describes how laws of physics transform actions into outcomes. If $u^i : Z \to \mathbb{R}$ denotes player i's utility function, then we assume that $v^i = u^i \circ P$.

Let $Z^\infty = \underset{t=1}{\overset{\infty}{\times}} Z_t$, where $Z_t = Z$ for all $t \in \mathbb{N}$. Let $\Omega^\infty := \Omega \times \Omega \times ...$ be the countable infinite Cartesian product of Ω, and $(\Omega^\infty, \mathcal{G}, \mu)$ denote the usual corresponding probability space (see for example, Fristedt and Gray (1997), chapter 9.) A generic element of Ω^∞ is denoted by $\omega^\infty = \{\omega_t\}_{t=1}^\infty$, where $\omega_t \in \Omega$, for all $t \in \mathbb{N}$. We define, for $\delta \in (0,1)$,

$$U^i(z^\infty, \omega^\infty) = (1-\delta)\sum_{k=1}^{\infty} \delta^{k-1} u^i_{\omega_k}(z_k),$$

With a slight abuse of notation, we write

$$\begin{aligned}U^i(\sigma) &= \int_{\Omega^\infty} U^i(z^\infty_\sigma(\omega^\infty), \omega^\infty) d\mu = \\ &= (1-\delta)\sum_{k=1}^{\infty} \delta^{k-1} v^i_k(\sigma),\end{aligned}$$

where,

$$u_k^i(\sigma) = \sum_{\omega^k \in \Omega^k} p(\omega_1) g(\omega_2|\omega_1, a_1(\omega^k)) \cdots g(\omega_k|\omega_{k-1}, a_{k-1}(\omega^k)) v_{\omega_k}^i(P(a_k(\omega^k))),$$

and $z_\sigma^\infty(\omega^\infty) = (P(a_1(\omega^1)), P(a_2(\omega^2)), \ldots)$.

3 Outcome Functions

At this point, we can establish some connections between our framework and the one defined in Hurwicz (1994). An outcome function is a function $\mathcal{M} : \Sigma \times \Omega^\infty \to Z^\infty$; it describes what will happen as a consequence of players' choices of strategies for each possible realization of the uncertainty.

Given an outcome function $\mathcal{M}$, we say that a strategy σ is a Nash equilibrium given $\mathcal{M}$ if for all $i \in N$, and all $\sigma_i' \in \Sigma_i$

$$U^i(\mathcal{M}(\sigma)) \geq U^i(\mathcal{M}(\sigma_i', \sigma_{-i})).$$

As it is clear from the definition, all we need to describe is the consequences of unilateral deviations. Thus, we define, for any given strategy σ the function $\mathcal{M}_\sigma^i : \Sigma_i \times \Omega^\infty \to Z^\infty$ by

$$\mathcal{M}_\sigma^i(\sigma_i', \omega^\infty) = \mathcal{M}((\sigma_i', \sigma_{-i}), \omega^\infty).$$

Note that

$$\mathcal{M}_\sigma^i(\sigma_i, \omega^\infty) = \mathcal{M}(\sigma, \omega^\infty);$$

and furthermore,

Remark 1 *A strategy σ is a Nash equilibrium given $\mathcal{M}$ if and only if, for all $i \in N$, and all $\sigma_i' \in \Sigma_i$*

$$U^i(\mathcal{M}_\sigma^i(\sigma_i)) \geq U^i(\mathcal{M}_\sigma^i(\sigma_i')).$$

That is, the knowledge of $\mathcal{M}_\sigma^i$ for all $\sigma \in \Sigma$ and $i \in N$, which we call *player i's partial outcome function induced by σ*, is all we need to discuss Nash equilibria. Similar considerations apply for subgame perfection.

Returning to the framework of the previous section, we recall that a strategy σ determines, for each realization of the uncertainty, a sequence of actions $a^\infty(\sigma, \omega^\infty)$: $a_1(\sigma, \omega^\infty) = \sigma(\omega_1)$ and

$$a_t(\sigma, \omega^\infty) = \sigma(\omega_1, a_1(\sigma, \omega^\infty), \ldots, \omega_{t-1}, a_{t-1}(\sigma, \omega^\infty), \omega_t).$$

For each $i \in N$, and $\sigma \in \Sigma$, we define $\mathcal{O}^i_\sigma : \Sigma_i \times \Omega^\infty \to Z^\infty$ by

$$\mathcal{O}^i_\sigma(\sigma'_i, \omega^\infty) = a^\infty((\sigma'_i, \sigma_{-i}), \omega^\infty).$$

The function $\mathcal{O}^i_\sigma$, which we call *player i's outcome function induced by* σ, gives the outcome that player i expects if he uses strategy σ_i, the realization of uncertainty is ω^∞ and he believes that everyone else is behaving according to σ.

Our objective is to study equilibrium strategies. We assert that although a player-specific outcome function induced by a given strategy does not correspond to an outcome function, it corresponds to a partial outcome function, which, is enough for equilibrium analysis.

Remark 2 *A strategy σ is a Nash equilibrium if and only if for all $i \in N$, and $\sigma'_i \in \Sigma_i$,*

$$U^i(\mathcal{O}^i_\sigma(\sigma_i)) \geq U^i(\mathcal{O}^i_\sigma(\sigma'_i)).$$

However, if players' preferences are selfish, that is, if for all $i \in N$ and $z, \bar{z}$ such that $z_i = \bar{z}_i$ we have $u_i(z) = u_i(\bar{z})$, we may associate an outcome function with a given strategy σ as follows. For $\sigma \in \Sigma$, and $i \in N$, let Ψ^i_σ denote the i^{th} projection of $\mathcal{O}^i_\sigma$. Define $\Psi_\sigma : \Sigma \times \Omega^\infty \to Z^\infty$ by

$$\Psi_\sigma(\sigma') = (\Psi^1_\sigma(\sigma'_1), \ldots, \Psi^n_\sigma(\sigma'_n)).$$

Since Ψ_σ maps the joint strategy space into the outcome space, we may call Ψ_σ the *outcome function induced by* I. It describes what outcome each player is expecting to get, when he believes that the others will follows σ.

Remark 3 *Suppose that for all $i \in N$, player i's preferences are selfish. Then, a strategy σ is a Nash equilibrium if and only if, for all $i \in N$, and $\sigma'_i \in \Sigma_i$,*

$$U^i(\Psi_\sigma(\sigma)) \geq U^i(\Psi_\sigma(\sigma'_i, \sigma_{-i})).$$

Finally, we define the notion of an equilibrium social institution. Since we defined a social institution as a family of strategies, let $\mathcal{S} \subseteq \Sigma$ denote a social institution. We say that $\mathcal{S}$ is a Nash (subgame perfect) equilibrium social institution if there exists $\sigma \in \mathcal{S}$ such that σ is a Nash (subgame perfect) equilibrium.

Here, it may help to recall an example we used in the introduction. Consider a legal code in a given society, which classifies certain actions of certain

people as 'illegal,' and prescribes a certain punishment, to be enforced by a different group of people. As before, we associate that legal code with the set of all strategies respecting the above properties, i.e., all the strategies that will trigger the corresponding punishment whenever an illegal action is taken. Thus, if two strategies respect the defining properties of a legal code but differ on the legal actions a given player takes, then we would still say that the legal code prevails in both cases. And if one one of these two strategies is an equilibrium, we still say that the legal code is an equilibrium, in the sense that it is plausible (in an equilibrium sense) that such a legal code endures.

4 Examples

4.1 A First Example

The following economic example illustrates our formulation.[3] There are three people in an island, two of whom are farmers and the other one is a guardian. One farmer lives in the North part of the island, while the other farmer lives in the South part. If the weather is good (for farming) in the South (resp. North,) then the Southern (resp. Northern) farmer can produce three loaves of bread. However, the meteorological conditions in this island are such that whenever the weather is good in the South part, the weather is bad in the North part, and vice versa. Finally, the guardian cannot produce any good at all, but he is strong enough to force each one of the others to do what he wants them to do. However, he is very "modest", in the sense that he will be happy provided that he can eat in every period.[4]

Let us formulate this story as a normal form game. The set of players is $N = \{1, 2, 3\}$, with the convention that player 1 is the Northern farmer, player 2 is the Southern farmer, and player 3 is the guardian. In every period there is exactly one farmer with bread. This farmer can chose to eat the three loaves of bread, but he can also give some to the others. Hence, the choice

[3]This example is a modified version of an example commonly attributed to J. Hirshleifer. See Murota (1976, Chapter 4).

[4]This last assumption is made to facilitate guarding the guardian. See Hurwicz (1998) for more on the "guarding the guardians" problem.

set for player $i = 1, 2$ is

$$A^i = \{(c^1, c^2, c^3) \in \{0, 1, 2, 3\}^3 : \sum_{j=1}^{3} c^j = 3\}.$$

Also, c_t^j is player j's consumption and also the gift from player i to player j, $j \neq i$ in period t.[5]

Regarding player 3, the guardian, he can kill player 1, or player 2, or he can force any allocation in $A^{i(t)}$, with $i(t) = 1$ if t is odd and $i(t) = 2$ if t is even. Also, he can choose not to do anything. His action set is then

$$A^3 = \{k^1, k^2, \{(c^1, c^2, c^3)\}_{(c^1,c^2,c^3)\in A^{i(t)}}, na\}$$

We assume that the weather is good in the South (resp. North) in every even (resp. odd) period. We also assume that the discount factor δ is 0.95.

The outcome space is

$$Z = \{(c^1, c^2, c^3) \in \{0, 1, 2, 3\}^3 : 0 \leq \sum_{j=1}^{3} c^j \leq 3\}.$$

The utility functions are

$$u^i(z) = \sqrt{z^i},$$

for the farmers, and

$$u^3(z) = \begin{cases} 1 \text{ if } z \geq 1 \\ 0 \text{ otherwise} \end{cases}$$

for the guardian.

For the physical outcome function we assume the following:

$$P(a_t^1, a_t^2, a_t^3) = \begin{cases} a_t^3 & \text{if} \quad a_t^3 \in A_t^{i(t)} \\ a_t^j & \text{if} \quad j = i(t) \text{ and } a_t^3 = na \\ (0, 0, 0) & \text{if} \quad a_l^3 = k^{i(t)} \text{ for some } l \leq t. \end{cases}$$

The first condition says that the guardian can force any allocation in $A_t^{i(t)}$ if he wishes. The second condition says that if the guardian does not intervene,

[5]Note that player i won't be able to provide any allocation in A^i if the weather is bad in his part of the island. We will take care of this problem with the physical outcome function.

then the resulting allocation is the one proposed by the productive farmer. Finally, the third condition says that if the productive farmer was killed in the past, then there is no output.

In this economy there are potential gains from a social institution that would promote gifts between the farmers, to avoid the uneven consumption path that both farmers would experience under autarky. Given our assumption on the discount factor, the gains due to consumption smoothing provided by such social institution are high enough to pay the guardian to enforce the institution.

In what follows, we will present several examples of social institutions that result $z_t = (1, 1, 1)$ in all periods. Since not all will be equilibrium social institutions, we will be able to conclude, by studying the incentives of all the members of this society, that some institutions are implementable while some other are not. Hence, the way an outcome is implemented also matters.

One possible social institution, denoted $\tilde{\sigma}$, can be described as follows: if the outcome in the past has always equal to $(1, 1, 1)$, then both farmers propose $(1, 1, 1)$ and the guardian does not do anything (i.e., he chooses na.) If player 1 has provoked a deviation from the outcome $(1, 1, 1)$, then the guardian kills him (even if he is already dead) and player 2 proposes $(0, 2, 1)$ (player 1's choice is irrelevant, and therefore omitted.) A similar definition holds for histories in which player 2 has provoked a deviation from the outcome $(1, 1, 1)$. Finally, player 3's deviations go unpunished, as well as multi-player deviations.

However, this strategy is not a Nash equilibrium. This follows because the guardian has an incentive not to kill a deviant farmer: if he doesn't kill he will receive one loaf of bread in every period, and so an utility of 1; if he kills he will receive one loaf of bread every other period, and so an utility of $\frac{1-\delta}{1-\delta^2} \simeq 0.5$.

Suppose we define another strategy σ^* simple changing the action taken by the guarding after a deviation by any farmer to $(1, 1, 1)$. That is, after a deviation from any farmer, the guardian forces the productive farmer to provide one bread to everyone. In this case, it is easy to verify that σ^* is a Nash (and subgame perfect) equilibrium.

We would like to use this simple example to make two points: the first is that the function P describes only aspects of physical nature, while the strategies ($\tilde{\sigma}$ or σ^*) describe the behavioral aspects of (a particular instance of) the society.

Before addressing the second point, note that in $\tilde{\sigma}$ player 1 receives the

following amounts of bread (i.e., we will describe the first component of the function $\mathcal{O}^1_{\tilde{\sigma}}$): he gets 1 in every period if he never deviates. If he deviates, then he gets 1 until he deviates, a^1 in the period he deviates (if he deviates in an even period he gets 1 instead), and 0 afterwards.

Suppose now that we describe this economy omitting the guardian and by adding the players specific outcome functions $\mathcal{O}^1_{\tilde{\sigma}}$ and $\mathcal{O}^2_{\tilde{\sigma}}$. In this case, we would conclude that the restriction of $\tilde{\sigma}$ is an equilibrium. However, when we add the guardian as a player, we see that he has an incentive not to carry the enforcement implicit in $\tilde{\sigma}$. Hence, we can conclude that the restriction of $\tilde{\sigma}$ is not genuinely implementable.

In conclusion, we have demonstrated that by making all actors explicit players in the game, we can check whether a social institution is genuinely implementable by studying whether it is an equilibrium of the enlarged game.

4.2 A Second Example

Here we present another example of the framework we have in mind. As before, the set of player is denoted by N, and A_i is the set of player i's actions, for all $i \in N$. Let $A = \times_{i \in N} A_i$. In this example we assume that there is no uncertainty, and that the interaction between the player takes place only in one period.[6] Let Z be the outcome space, and $P : A \to Z$ be the physical outcome function; the function P describes only physical aspects.

Let us assume that we can partition the set of player into the set I of private agents, and the set $I^c = N \backslash I$ of public agents. The latest set includes, for example, congressmen, policemen, bureaucrats, etc. We will interpret different actions by the public agents to correspond to different policies or different laws; in the context of this example we use 'rules' synonymously with 'actions by the public agents.' Corresponding to the above partition of players, we let $A_I = \times_{i \in I} A_i$, and $A_{-I} = \times_{i \in I^c} A_i$.

A vector of actions for the public agents is $a_{-I} \in A_{-I}$. For each a_{-I}, we define a function $\mathcal{O}_{a_{-I}} : A_I \to Z$ by

$$\mathcal{O}_{a_{-I}}(a_I) = P(a_I, a_{-I}).$$

The function $\mathcal{O}_{a_{-I}}$ is a partial outcome function, mapping the actions that

[6] Alternatively, we may think of the elements of A_i as being denumerable vectors, each of the coordinates representing an action taken in a given time period.

private agents take into consequences — we refer to it as the private agents' partial outcome function.

The private agents' partial outcome function implicitly describes rules that private agents face, and which will influence their behavior. Given our interpretation that the function P describes only physical aspects, then the behavior aspects of the rules will have to be determined by the vector a_{-I}. Hence, we can describe any rule in this economy either by a strategy for the public agents, or by the private agents' partial outcome function that it induces. Alternatively, we may describe a rule by a family of strategies $\mathcal{R}_{a_{-I}} = \{\tilde{a} \in A : \tilde{a}_{-I} = a_{-I}\}$.[7]

One advantage of this formulation is that it allows us to study the incentives of the agents that formulate, and execute, the rules — in particular, we can ask whether the agents that are responsible for the enforcement actually have an incentive to enforce the rule. Also, we can ask whether a given rule will be approved in the congress, etc.

5 Conclusion

We proposed to define social institutions as families of strategies in some stochastic game. Social institutions were already associated with strategies by Schotter (1981) and Okuno-Fujiwara and Postlewaite (1995); furthermore, they were defined as families of game forms by Hurwicz (1994) and Hurwicz (1996). Since strategies induce game forms (in a particular sense developed here,) we conclude that our definition of social institutions naturally reflects the previous, apparently conflicting, definitions.

References

Fristedt, B., and L. Gray (1997): *A Modern Approach to Probability Theory*. Birkhäuser, Boston.

Hurwicz, L. (1994): "Institutional Change and the Theory of Mechanism Design," *Academia Economic Papers*, 22, 1–27.

[7] In this paragraph, we implicitly assume that any rule can be associated uniquely with a vector of actions for the public agents. More generally, a rule may be associated with a family of actions for the public agents, which will induce a family of private agents' partial outcome functions.

——— (1996): "Institutions as Families of Game Forms," *The Japanese Economic Review*, 47, 113–132.

——— (1998): "But Who will Guard the Guardians?," *mimeo*.

Murota, T. (1976): "Public Information and Social Welfare under Five Alternative Market Mechanisms," Ph.D. thesis, University of Minnesota.

North, D. (1990): *Institutions, Institutional Change and Economic Performance*. Cambridge University Press, Cambridge.

Okuno-Fujiwara, M., and A. Postlewaite (1995): "Social Norms and Random Matching Games," *Games and Economic Behavior*, 9, 79–109.

Schotter, A. (1981): *The Economic Theory of Social Institutions*. Cambridge University Press, Cambridge.